REQUEST FOR A SPECIAL PROJECT 2025–2027

MEMBER STATE: Germany

Principal Investigator[1]**:** Prof. George Craig

Affiliation: Ludwig-Maximilians-Universität München

Address: Theresienstr. 37, 80333 München, Germany

Other researchers: Dr. Tobias Selz

Project Title: Flow-dependence of the intrinsic predictability limit and its relevance to forecast busts

To make changes to an existing project please submit an amended version of the original form.)

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SPDECRAI**	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2025	
Would you accept support for 1 year only, if necessary?	YES ☐	NO ☒

Computer resources required for project year:		**2025**	**2026**	**2027**
High Performance Computing Facility	[SBU]	120M	120M	120M
Accumulated data storage (total archive volume)[2]	[GB]	0	0	0

EWC resources required for project year:		**2025**	**2026**	**2027**
Number of vCPUs	[#]			
Total memory	[GB]			
Storage	[GB]			
Number of vGPUs[3]	[#]			

Continue overleaf.

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide annual progress reports of the project's activities, etc.
[2] These figures refer to data archived in ECFS and MARS. If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year etc.
[3] The number of vGPU is referred to the equivalent number of virtualized vGPUs with 8GB memory.

Principal Investigator: Prof. George Craig

Project Title: Flow-dependence of the intrinsic predictability limit and its relevance to forecast busts

Extended abstract

Starting point

Despite the constant progress in weather prediction, significant dropouts in forecast quality continue to occur. Although quite rare, these so-called forecast busts may cause large negative impacts to society and economy, in particular if they are also associated with high-impact weather. Figure 1 illustrates a recent occurrence of a very severe forecast bust at the end of August 2023 over Europe (forecast init time). Not only ECMWF, but also other weather centers suffered from exceptionally low forecast skill, including the newly-developed AI-based models. Although the frequency of forecast busts has been reduced over the past decades (Lillo and Parsons, 2017), the question arises whether at least some of them are a consequence of a transiently low intrinsic predictability limit and even with almost perfect forecasting methods it will be impossible to avoid them.

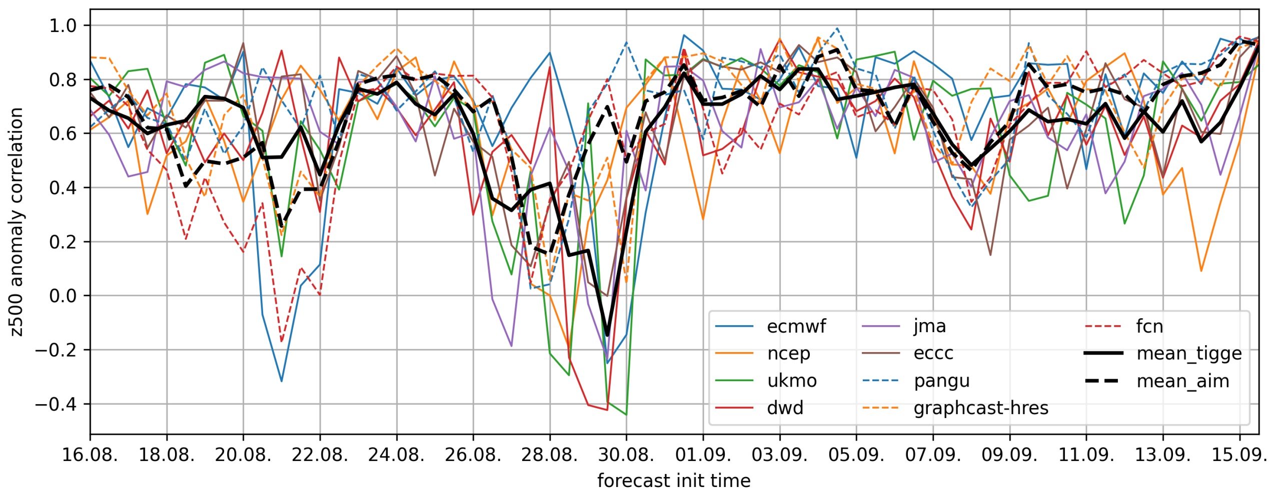

Figure 1: Anomaly correlation of the 500hPa geopotential over Europe at 7 days lead time for one month of initialization times (x-axis, 2023). Included are 5 operational models from the TIGGE archive and 3 recent artificial-intelligence based models. The period includes a very pronounced forecast bust ranging from 26 Aug to 30 Aug (init time).

Forecast busts over Europe were first studied by Rodwell et al., 2013 and have later been sorted into categories by Lillo and Parsons, 2017. Particularly interesting are: Extra-tropical transition of tropical cyclones (ET) and strong mesoscale convective systems (MCS) over the North American continent around the forecast initial time. Extra-tropical transition is a known cause of low practical predictability due to uncertainties in the cyclone track prediction and high sensitivities to the phasing when the cyclone enters the midlatiude wave guide (Keller et al., 2019). On the other hand, convection over North America can significantly modify upper tropospheric dynamics by negative potential vorticity generation and vertical momentum transport. Both of these scenarios directly couple uncertain small-scale, non-conservative processes to the large-scale dynamics downstream.

In general, the atmospheric circulation is difficult to predict. Its chaotic and non-linear evolution leads to exponential growth of the unavoidable initial condition uncertainties. On top of that, the atmosphere consists of many scales of motions ranging from turbulent eddies in the boundary layer through shallow to deep convection, to cyclones and Rossby waves. Due to interactions between these scales, initial perturbations of tiny amplitude lead to a cascade-like upscale error growth and eventually to a fundamental intrinsic limit of predictability, which cannot be overcome, even with perfect observations and models. This phenomenon is known as the "butterfly effect" (Lorenz 1969, Palmer et al., 2014). Research has shown that compared to current forecasting capabilities this intrinsic limit is still 4-5 days away in the midlatitudes (Zhang et al., 2019, Selz et al., 2022) for average atmospheric conditions. However, like operational predictability, it can be expected that the intrinsic predictability limit and hence the improvable forecast range can vary greatly, both locally and from case to case. Research suggests that the practical and intrinsic limits are not necessarily related, since the physical processes responsible for initial uncertainty growth are different, depending on the amplitude of the initial uncertainty (Selz et al., 2022). In the current special project, we are investigating the spatial-temporal variability of the intrinsic limit and the improvement potential systematically by running ensemble forecasts from varying levels of initial condition uncertainty for a large number of cases (see progress report).

The types of bust cases described above largely involve diabatic processes on relatively small scales that are difficult to model and hence prone to errors and biases. By correcting such errors, a significant reduction in the severity and frequency of the busts can be expected. On the other hand, small-scale diabatic processes have also been linked to fast upscale growth of tiny, butterfly-like perturbations and play a crucial role in setting an intrinsic limit. Therefore, it seems reasonable to hypothesize, that some fraction of the forecast busts could be a consequence of a very short intrinsic limit under certain rare conditions. In this case, improvements of the forecast busts cannot be expected, or only in a probabilistic sense.

Recently, an unintended butterfly experiment has been started by ECMWF (Linus Maggnuson, personal communication). With the resolution upgrade in Cycle 48r1, implemented on 27 June 2023, the deterministic IFS forecast and the unperturbed ensemble control forecast should produce bit-identical forecasts. However, they use different orography files, which introduce tiny, butterfly-like perturbation after the first time step. Figure 2 compares the spread of this two-member butterfly ensemble to the spread of the 50 member ensemble prediction system (IFS-ENS) at 7-days forecast lead time over Europe (red and blue line, respectively). The large average difference between these curves indicates the remaining improvement potential. However, there is one incident (init time 28 Aug 2023, 0UTC) where the spread of the butterfly ensemble becomes equal to the spread of the 50 member ensemble operational ensemble. Interestingly, this date is related to the bust period mentioned earlier (Figure 1). Although there are still many open questions and the result is largely limited by sampling uncertainty, it clearly suggests that cases may exist, where our current forecast capabilities have already hit the intrinsic limit.

An ongoing analysis of this case has revealed that it involves an ET event early in the forecast combined with a block onset over Europe around the evaluation time. By chance, the ICON model ensemble experiments conducted during the ongoing special project (see progress report) also include this case, however they are giving conflicting results (dots in Figure 2). The uncertainty after 7 days of the "butterfly" simulations (here started from a EDA sample reduced to 10%) remained about average, while the spread of the 100%-EDA ensemble strongly increased. This would indicate that the bust is not caused by a short intrinsic limit but possesses an unusual large improvement potential. A preliminary investigation of the simulations showed a systematic bias in ICON in the evaluation of the tropical cyclone position, which may lead to a miss of the sensitivity or bifurcation point. Interestingly, the IFS-butterfly ensemble only shows exceptionally large spread for a single initialization time, while the associated forecast bust lasted 2-3 days. It is plausible that for the other init. times IFS could have missed the bifurcation point as well. This observation adds another level of complexity to the investigation of forecast busts.

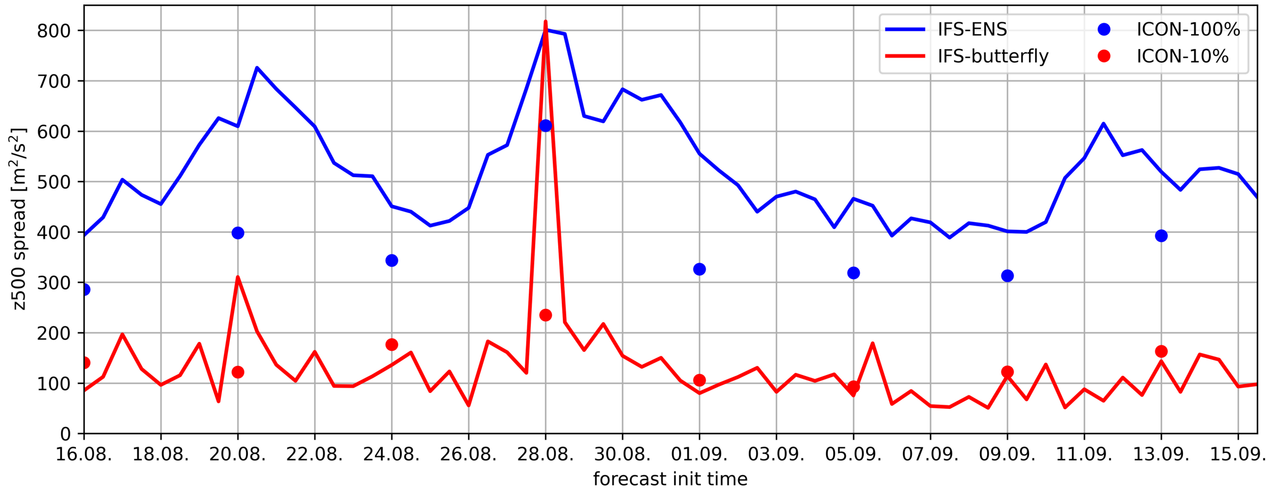

Figure 2: 500hPa geopotential spread over Europe at 7 days forecast lead time for one month of initialization times (2023) and for 4 different ensembles. Blue colors represent ensembles (IFS and ICON), started from current initial condition uncertainties. Red colors represent ensembles, started from small, "butterfly-like" uncertainty amplitudes.

The second atmospheric configuration associated with forecast busts features strong mesoscale convective systems (MCS) over North America early in the forecast. Such situations usually occur in spring and summer. With this year's computational resources, we are working on investigating these types of busts and their improvement potential using rather low-resolution simulations together with a stochastic convection scheme to account for the missing variability from upscale error growth in convection. Results from this research is expected by next year to provide a partial answer to this question. There are however limitations to this approach, especially missing or insufficient convective organization and potential biases in convective outflow level and the vertical momentum transport from the convection scheme. Given those issues, the low-resolution simulations might miss the crucial sensitivity and suggest an incorrect conclusion.

Scientific plan

Given these open but very relevant questions, we apply for a continuation of our special project to further research uncertainty growth in forecast bust cases, in particular with respect to small amplitudes of initial condition uncertainty. This will hopefully further clarify the significance of the butterfly effect for operational forecasting and provide an answer to what amount frequency and severity of forecast busts could be reduced with future technological improvements, including those based artificial-intelligence. Our plan to answer these questions is a detailed investigation of 10 past bust cases (5 of the ET-type and 5 of the MCS-type) through generating ensembles started from very small initial condition uncertainty. We will study the uncertainty growth in these ensembles and compare it to uncertainty growth of operational ensembles (IFS-ENS) to estimate how close the intrinsic limit is and how much room for improvement is present.

First, a list of bust cases is required, categorized into ET-type busts and MCS-type busts. To achieve this, we will analyze at least 10 years of forecasts from IFS-DET, IFS-ENS, and from the TIGGE archive. Already within the ongoing project we have obtained most of the relevant data from ECMWF's MARS archive. A first indication of a short intrinsic limit causing the bust would be that all models produce bad forecasts, although the extent and timing may vary (e.g. Figure 1). This analysis will be complemented by a sample of AI-based forecasts, and for recent initialization times by results from the IFS “butterfly” experiment. New results from the ongoing special project will also be considered. Though the focus of the bust cases selection will be Europe, we will also look into other parts of the world. While the MCS-type mechanism might be specific to the land-sea configuration upstream of Europe, we expect that ET-type bust cases could occur downstream of any region with tropical cyclone activity.

The two types of busts pose different challenges to numerical simulation. As shown in the introduction, uncertainty growth of ET-type cases can be highly sensitive to the basic state and whether or not the ensemble is centered around a bifurcation point. Missing the bifurcation point due to errors or biases in the model or in the initial condition estimate can lead to slow uncertainty growth and to an underestimation of the (intrinsic) predictability. To circumvent this problem, we plan to sample not only the small “butterfly-like” initial condition uncertainty, but also sample different basic states, where the ensemble is centered around. Hence for every bust case, we consider not only one analyses, but 5, taken from the EDA initial condition sample. In addition, 4 consecutive initialization times (each 12h apart) will be considered. From each of these 20 analyses, a 20-member ensemble will be computed, now sampled from small-amplitude uncertainty (e.g. 1% of the EDA spread or small-amplitude noise). To also account for model-specific biases each ensemble will be carried out with 2 different models (ICON and IFS) close to their operational resolution. We hypothesize that if the bust was caused by a low intrinsic limit, some of the ensembles will be centered around a bifurcation point causing the spread of the initially close 20 members to increase rapidly and after 5-7 days of lead time reach a level comparable to the spread of the full operational ensemble and comparable to the size of the error.

With the MCS-type bust cases, we face a different challenge. Here, accurate representation of small-scale convective processes, divergent outflow levels, vertical momentum transport and convective organization is crucial. On the other hand, we expect bifurcation points, analysis uncertainty or ensemble size to be less of an issue. Hence for this type of bust we will perform convection-permitting simulations with ICON (2.5km grid size), which as a non-hydrostatic model is more suited for these kinds of simulations than IFS. Due to the high computational cost, we plan to use “butterfly”-perturbation ensembles of 5 members centered around the analysis and to consider 2 initialization times for each case (12h or 24h apart). These runs will be computed up to 3 days lead time and after that continued at lower resolution for another 7 days, which will save a significant amount of computing time. We don’t expect any disadvantages from this approach, since the initially developing convective scale uncertainties will have amplified and grown upscale by 3 days, and a lower resolution will be sufficient to propagate them further in time (Selz et al. 2022).

Computing time estimates

We plan to run numerical simulations using two different models: IFS and ICON. Both are FORTRAN-based, MPI-parallelized models and are suitable for running on the Atos computer. With ICON we have lots of experience doing so, even at convection-permitting resolution. An estimate of the required computing time is given below, which is based on past ICON simulations. We assume that IFS at similar resolution requires a similar amount of computing time.

	ET-type busts	MCS-type busts
Number of cases	5	5
Number of analyses	5	1
Number of init times	4	2
Number of models	2	1
Ensemble size	20	5
Forecast lead time	8 days	3 days[2]
total number of simulations	4,000	50
cost per simulation [SBU]	23,000	4,400,000
Total cost [SBU]	92,000,000	220,000,000

For testing, errors and unexpected problems we add a buffer of 20% to the computing time, leading to a request of 120 MSBU per year. However, the exact experimental design will be adjusted based on incoming results from the ongoing special project, new publications and results from first tests and experiments. The design can also be adjusted if the requested large amount of computing time is not available, e.g. by reducing the number of cases that are considered.

References

Keller, J. H., and Coauthors, 2019: The Extratropical Transition of Tropical Cyclones. Part II: Interaction with the Midlatitude Flow, Downstream Impacts, and Implications for Predictability. Mon. Wea. Rev., 147, 1077–1106, https://doi.org/10.1175/MWR-D-17-0329.1.

Lillo, S. P., and D. B. Parsons, 2017: Investigating the dynamics of error growth in ECMWF medium-range forecast busts. Q. J. R. Meteorol. Soc., 143: 1211-1226, https://doi.org/10.1002/qj.2938.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. Tellus 21.3, 289-307, https://doi.org/10.3402/tellusa.v21i3.10086.

Palmer, T. N., A. Döring, and G. Seregin, 2014: The real butterfly effect. Nonlinearity 27.9, R123, https://doi.org/10.1088/0951-7715/27/9/R123.

Rodwell, M. J., and Coauthors, 2013: Characteristics of Occasional Poor Medium-Range Weather Forecasts for Europe. Bull. Amer. Meteor. Soc., 94, 1393–1405, https://doi.org/10.1175/BAMS-D-12-00099.1.

Selz, T., M. Riemer, and G. C. Craig, 2022: The Transition from Practical to Intrinsic Predictability of Midlatitude Weather. J. Atmos. Sci., 79, 2013–2030, https://doi.org/10.1175/JAS-D-21-0271.1.

[2] Continued with lower resolution up to 10 days, which only adds insignificant cost.